Structure of a *Neurospora* RNA polymerase I promoter defined by transcription *in vitro* with homologous extracts

Brett M. Tyler* and Norman H. Giles

Department of Genetics, University of Georgia, Athens, GA 30602, USA

Received 2 April 1985; Revised and Accepted 31 May 1985

ABSTRACT

A *Neurospora* *in vitro* transcription system has been developed which specifically and efficiently initiates transcription of a cloned *Neurospora crassa* ribosomal RNA gene by RNA polymerase I. The initiation site of transcription (both *in vitro* and *in vivo*) appears to be located about 850 bp from the 5' end of mature 17S rRNA. However, the primary rRNA transcripts are normally cleaved very rapidly at a site 120-125 nt from the 5' end *in vitro* and *in vivo*. The nucleotide sequence surrounding the initiation site has been determined. The region from -16 to +9 exhibits partial homology to the corresponding sequences from a wide variety of organisms including yeast, but the most striking similarity is to the initiation region from *Dictyostelium discoideum* which displays 73% homology to the *Neurospora* sequence from -23 to +47. The *Neurospora* sequences from -96 to +97 have been shown to be sufficient for transcription. This region contains two sequences displaying 8/9 bp matches to elements of the 5S rDNA promoter.

INTRODUCTION

In eukaryotes, RNA polymerase I is responsible for synthesis of the large ribosomal RNA precursor which yields the large rRNA's and 5.8S rRNA. The other components of the ribosome are encoded by genes transcribed by RNA polymerase II (the ribosomal proteins) and RNA polymerase III (5S rRNA). The mechanism by which the activities of these three transcriptional systems are regulated so as to achieve a co-ordinated supply of ribosomal components is poorly understood. In a variety of organisms, systems have been developed which faithfully initiate RNA polymerase I transcription *in vitro* (1-6). These systems have been used to show that specific initiation requires the participation of several protein factors in addition to purified RNA polymerase I (7,8), and that DNA sequences both upstream and downstream of the initiation site are required (9-11). Similar results have been obtained with RNA polymerase II and RNA polymerase III *in vitro* transcription systems (12-16).

Neurospora crassa is a useful organism for examining the role of gene regulation in ribosome biogenesis since it has a well-studied genetic system the large rRNA genes (17) and 5S rRNA genes (18) have been cloned, and kRNA polymerase II (19) and RNA polymerase III (20) in vitro systems have been developed. We report here the development of an accurate and efficient RNA polymerase I transcription system from Neurospora. We have used the system to localize the initiation site of rRNA transcription and to show that the 5' end of the primary transcript is rapidly processed. We have determined the nucleotide sequence of the DNA surrounding the initiation site and have localized the region required for transcription initiation.

MATERIALS and METHODS

Strains and Plasmids

The N. crassa strain normally used for transcription extracts was 105C (A qa-1S^C) (20). The rDNA template plasmids were derived from pKD2 (17) which contained the 5' half of the rDNA repeat on a 4.8 kb HindIII fragment. pKD2 was a generous gift of S. Free and K. Metzenberg. Portions of pKD2 were subcloned into pBR322, pUC9 (21) or pUC12 (21). The hybrid 5S rDNA/large rDNA template used in the experiment shown in figure 3 was created by removing the termination signal of a 5S rRNA gene with Bal31 exonuclease and inserting the gene between the SalI and BamHI sites of pBR322 (The cloned 5S rRNA gene (18) was a gift of E. Selker and K. Metzenberg). The DraI-EcoRI fragment containing the rDNA promoter region was then placed downstream of the 5S rRNA gene between the EcoRV and EcoRI sites of the plasmid. To construct the inverted repeat template used in the experiment of figure 4 one copy of the promoter region on a HindIII-PstI fragment was inserted into the HindIII and PstI sites of pUC9. The other copy on a BamHI - PvuII fragment (containing plasmid DNA at the BamHI (5') end) was then inserted into the SalI site (blunted with T₄ DNA polymerase) and the BamHI site in the polylinker of the vector. The resultant plasmid contained inverted repeats of the 1.3 kb HindIII - PstI portion of the rDNA promoter region separated by a 90 bp spacer. This plasmid was very stable and replicated prolifically in JM103 (21).

Isolation of Transcription Extracts

Transcription extracts were isolated exactly as described (19,20). In outline, macroconidia were germinated for 4 hours, harvested and then disrupted in a 'bead beater' (Biospec products). The lysate was made 0.9M in $(\text{NH}_4)_2\text{SO}_4$ then centrifuged at 180 000 g for 2 hr. Proteins were precipitated

from the supernatant with 65% saturated $(\text{NH}_4)_2\text{SO}_4$, redissolved in a very small volume and then dialysed. Extracts made this way contained specific RNA polymerase I, RNA polymerase II and rRNA polymerase III initiation activity. To obtain extracts competent only in RNA polymerase I initiation activity the $(\text{NH}_4)_2\text{SO}_4$ concentration used to treat the initial lysate was increased to 1.5M. To obtain extracts competent only in RNA polymerase III initiation activity, the $(\text{NH}_4)_2\text{SO}_4$ concentration used to precipitate the extract was reduced to 50% saturation.

Transcription Reactions

Reactions (25 μl) typically contained 2-6 μl extract, 10 mM K HEPES pH 7.9 5 mM K_2EGTA (total $[\text{K}^+] = 17\text{ mM}$) 10 mM $\text{Mg}(\text{OAc})_2$ 1-2% glycerol (from the extract), 2.5 mM DTT, 400 μM each GTP, CTP and ATP, 50 μM UTP, 4 mM potassium phosphoenol pyruvate 10-100 ng linearised plasmid DNA and 1.0 μCi $[\alpha\text{-}^{32}\text{P}]$ UTP. Reactions were incubated 20 min at 30°C and the RNA extracted as described (20). For direct analysis the RNA was resolved by electrophoresis through 5-6% acrylamide/7 M urea gels (21) at 55°C . For analysis of snap-back transcripts, the RNA pellet was dissolved in 100 μl RNase buffer (0.3 M NaOAc, 10mM Tris.HCl pH 7.4, 10 mM Na_3EDTA) and digested with 10 μg RNase A at 30°C for 30-60 min 100 μl more RNase buffer containing 1% SDS 10 μg *E. coli* RNA and 20 μg proteinase K was then added and digestion continued at 37°C for 15-30 min. Digestion products were precipitated with 500 μl ethanol, dissolved in 10 μl formamide/dye/1% SDS and electrophoresed as described above. Transcription products were detected by autoradiography and their sizes determined by reference to DNA standards.

To quantitate transcription, specific transcripts were localized by autoradiography and excised from the gel. The gel slices were then counted by Cerenkov counting in water. To provide a background an identical sized gel slice was excised from the gel immediately above the region of each specific transcript.

5' End Mapping of Transcripts and in vivo RNA

To isolate specific transcripts, the transcription reaction was scaled up four-fold and the $[\alpha\text{-}^{32}\text{P}]$ UTP reduced to 0.25 μCi Reactions products were fractionated as usual by electrophoresis on 0.2 mm sequencing gels. Specific transcripts were located by autoradiography excised from the gel and eluted from the gel slices by soaking in 1 ml 0.5 M NaCl, 10 mM Tris.HCl pH 7.4, 1 mM Na_3EDTA 0.2% N-Lauryl-Sarcosine (Sarkosyl) overnight with shaking at room temperature. The eluted RNA was precipitated by addition of 5 μg *E. coli* RNA and 2.5 volumes of ethanol. To provide control RNA to monitor detection of

transcripts endogenous to the extracts, the preparative reaction was carried out in the absence of template DNA, and the reaction products fractionated by electrophoresis. If appropriate, the products were first treated with RNase. By reference to an adjacent track containing the relevant specific in vitro transcript, an appropriate size fraction of the endogenous RNA was cut from the gel and eluted. 5' end mapping of the recovered in vitro transcripts and of Neurospora RNA was as described (22). Neurospora RNA was isolated as described (22).

DNA Sequencing

DNA fragments were labelled at the 5' positions using calf alkaline phosphatase and polynucleotide kinase (23). Where necessary, 3' overhangs were converted to 5' overhangs using the 3' exonuclease activity of T_4 DNA polymerase. Labelled fragments were purified and sequenced by chemical degradation (23).

RESULTS

Specific Transcription of Neurospora rDNA

The 5' end of mature Neurospora 17S rRNA has previously been shown to be located near the unique XbaI site in the 9 kb rDNA repeat (17). In order to assay crude Neurospora extracts for the ability to initiate specific RNA polymerase I transcription, the 1.65 kb region upstream of the XbaI site was subcloned for use as a transcription template. Figure 1A shows a restriction map of this region. Template DNA was digested at a variety of positions with restriction endonucleases and incubated with a crude extract previously shown to specifically transcribe Neurospora 5S rRNA and tRNA genes (20). Figure 1B shows that when the template DNA was cleaved with XbaI, a predominant 700 nt run off transcript was produced (track a), suggesting that there was a specific RNA polymerase I initiation site 700 bp upstream from the XbaI site. When the template was cleaved closer to this putative initiation site with EcoRI, PvuII or PstI, shorter transcripts of 410 nt (track b), 290 nt (track c) and 190 nt (track d) respectively were produced. In each case, the length of the shorter transcript corresponded closely to the distance from the cleavage site to the putative initiation site defined by transcription of the XbaI template, indicating that the transcripts originated from a common site within the rDNA 5' region and terminated by run off at the restriction enzyme cleavage site. Transcription depended on template rDNA since no transcripts were produced from pUC9 plasmid DNA (21) (figure 1C, track b) and was inhibited by 4 mM cordycepin 5' triphosphate (track c), confirming that it was

Figure 1: Specific Transcription of Cloned *Neurospora* rDNA. (A) Restriction map of *Neurospora* rDNA 5' to the 17S rRNA coding region. Restriction enzyme sites are: Hd, HindIII; Dr, DraI; Xh, XhoI; St, SatII; Ps, PstI; Pv, PvuII; Rl, EcoRI; Xb, XbaI; X_m, XmnI. The structure of a subclone (pRRH/R) terminating at the EcoRI site and used as a template in figure 1C, track e is also shown. The vertical arrow indicates the 5' ends of the run-off transcripts. T₅ and T₆ indicate TTTT and TTTTTT sequences respectively which act as terminators of RNA polymerase III transcription. (B) Transcription of rDNA templates cleaved with different enzymes. Standard reactions (see materials and methods) contained 30 ng of pRRH/X (HindIII-XbaI subclone) cleaved with: track a, XbaI; track b, EcoRI; track c, PvuII; track d, PstI. nt = nucleotides. (C) Characterization of the transcription reaction. Standard reactions contained 30 ng EcoRI-cleaved pRRH/X DNA (tracks a, c, d), 30 ng EcoRI-cleaved pUC13 plasmid DNA (track b) or 30 ng XmnI-cleaved pRRH/R (HindIII-EcoRI subclone) DNA (track e). The reactions shown in tracks c and d also included 1 mM cordycepin 5' triphosphate or 1 mg/ml α -amanitin, respectively.

due to an RNA polymerase (24), however it was not inhibited by 1 mg/ml α -amanitin (track d), indicating that RNA polymerase II was not responsible for the transcripts (19). To rule out RNA polymerase III, which is also resistant to α -amanitin (20), as a source of the transcripts, the template sequences downstream of the EcoRI site were replaced with pBK322 plasmid sequences containing one T₆ and two T₅ sequences (figure 1A). T₅, and especially T₆ sequences act as efficient terminators of RNA polymerase III transcription in Xenopus (25) and in Neurospora (B. Tyler, unpublished). Figure 1C (track e) shows that transcription from the putative rDNA initiation site proceeds through all three potential termination sites to an XmnI cleavage site in the plasmid sequences of the template, producing an 840 nt transcript and ruling out RNA polymerase III as the source of the transcripts. Transcription therefore appears to be due to RNA polymerase I.

Properties of the Transcription Reaction

The crude Neurospora extracts used for the RNA polymerase I transcription system were identical to those used for Neurospora RNA polymerase II (19) and RNA polymerase III (20) transcription systems and many of the comments regarding these systems are also true of the RNA polymerase I system. However, changes in the salt concentration used for extraction of the lysate (optimally 0.9M (NH₄)₂SO₄), and in the ammonium sulphate concentration used to precipitate the extract (65% saturation) differentially affected the three systems. For instance, increasing the (NH₄)₂SO₄ concentration during extraction to 1.2 M or 1.5 M had no effect on RNA polymerase I transcription, whereas 1.2 M salt extraction abolished RNA polymerase II transcription and 1.5 M salt extraction additionally abolished RNA polymerase III transcription. Similarly, reducing the (NH₄)₂SO₄ concentration to 50% saturation at the final precipitation step reduced specific RNA polymerase I and II transcription more than 10 fold while specific RNA polymerase III transcription was unaffected. These differential effects of extraction conditions on the different transcription systems, combined with differences in optimal reaction conditions readily enabled the development of individual 'mono-specific' RNA polymerase I and RNA polymerase III transcription systems which proved useful in analysing transcription and processing events in these systems (see below).

The optimum temperature for the RNA polymerase I reaction is 30°C while the optimum extract concentration was 2-4 μ l per 25 μ l reaction. The optimum Mg²⁺ concentration was 10mM, as shown in figure 2A, while addition of any salt, especially ammonium sulphate, inhibited transcription (figure 2B).

Figure 2: Properties of the Transcription Reaction. Standard reactions containing 5 μCi [α - ^{32}P] UTP were used with the indicated variations. Transcription was quantitated by Cerenkov counting of gel slices. (A) Mg^{2+} concentration was varied by addition of $Mg(oAc)_2$. (B) Salt concentration was varied by addition of KOAc (filled circles, solid line) or $(NH_4)_2SO_4$ (open circles, broken line). (C) Variation of the template DNA concentration (Pst-cleaved) in the presence (open circles, broken line) or absence (filled circles, solid line) of circular pUC9 carrier DNA (added to adjust the total DNA concentration to 4.0 $\mu g/ml$). (D) Variation of the total DNA concentration. Varying amounts of carrier DNA were mixed with a constant amount (0.12 $\mu g/ml$) of template DNA.

The template DNA optimum for the RNA polymerase I reaction is very broad, with at least 70% maximal activity being obtained over a 100 fold range (0.4–40 $\mu g/ml$) (figure 2C). However, as shown by the broken line in figure 2C, when the DNA concentration was kept constant by adding carrier DNA the template DNA optimum is somewhat sharper. In fact, there was less transcription in the presence of carrier DNA than without it. Figure 2D confirms that the carrier DNA inhibited transcription, reducing it to 5% maximal at 400 $\mu g/ml$. Since the carrier DNA consisted of the same plasmid (pUC9) which was the vector for the rDNA template DNA, the broad template DNA optimum observed in the absence of carrier DNA appears to result from co-ordinate and competing increases (or decreases) in substrate and inhibitor concentrations. A consequence of the self-inhibition by the template DNA is

that the highest template efficiencies (20-30 transcripts/gene/20 min reaction) were obtained at the lowest template DNA concentrations (less than 0.1 µg/ml) whereas at template concentrations which gave maximum specific synthesis (for a given amount of extract) the template efficiency was relatively low (1-3 transcripts/gene/20 min).

Primary rDNA Transcripts are Rapidly Processed Near their 5' End

In addition to the run-off transcripts, a prominent 120 nt RNA was transcribed from the rDNA (not present on the gel shown in figure 1). The origin of this transcript was localized to the region between the XhoI and SstII sites, close to the putative RNA polymerase I initiation site (figure 1A) because, as shown in figure 3A, it was generated from subcloned rDNA templates terminating at the PstI (track a) and SstII (track b) sites, but not from one terminating at the XhoI site (track c).

To determine if the 120nt RNA might be cleaved from the 5' end of the primary rRNA transcript, a 5S rRNA gene devoid of its terminator sequence (B. Tyler, unpublished) was placed 400 bp upstream of the potential cleavage site (at the 5' end of the run-off transcripts) so that RNA polymerase III transcripts initiated from this gene would span the site. Plasmid sequences containing RNA polymerase III termination signals were placed downstream of the rDNA so that an unprocessed RNA polymerase III transcript initiated at the 5S rRNA gene would be predominantly 960 nt in length. If RNA cleavage occurred at a site corresponding to the 5' end of the rDNA run-off transcripts, then the RNA polymerase III transcript should be cleaved into upstream and downstream fragments 520 nt and 440 nt long, respectively. Figure 4C summarizes the structure of this hybrid template. Figure 3B, tracks a and b show that when this template was incubated with an RNA polymerase III transcription extract devoid of RNA polymerase I activity, two predominant sets of transcripts of around 520 nt and 440 nt were produced, with no trace of a 960 nt transcript detectable. When the template was cleaved with PvuII, upstream of the terminators (figure 3C), the upstream 520 nt RNA fragment was unaltered but the downstream set of fragments was reduced by runoff to around 290 nt (track c). Track e shows that the set of downstream run-off fragments produced from the PvuII-cleaved template by a specific RNA polymerase I extract was identical to that produced by the RNA polymerase III extract (track c), indicating that the cleavage reaction carried out by the two extracts was identical. Track d confirms that the extract used in track c was inactive on a conventional rDNA template. These data strongly suggest that the 5' ends of the run off transcripts are generated by a rapid and efficient processing event.

Figure 3: Evidence for Cleavage of the Primary Transcript. (A) Localization of the 120 nt transcript. Template DNAs (1.2 $\mu\text{g/ml}$) in standard reactions consisted of a HindIII-PstI (track a), a HindIII-SstII (track b) or a HindIII-XhoI (track c) subclone cleaved with PstI, SstII or XhoI respectively. nt = nucleotides. (B) Cleavage of RNA polymerase III read-through transcripts. The RNA polymerase I reaction (track e) was standard. The RNA polymerase III reactions (tracks a-d) resembled the RNA polymerase I reactions except that they contained 24 $\mu\text{g/ml}$ template DNA, 0.25 μCi [α - ^{32}P] UTP and 100 mM KOAc (24). Transcription extracts were specific for RNA polymerase III initiation (tracks a-d) or RNA polymerase I initiation (track e) activity. Template DNAs were the hybrid 5S rDNA-large rDNA plasmid either uncleaved (track a) or digested with XmnI (track b) or PvuII (tracks c and e). Track d contained a standard PvuII-cleaved rDNA template. nt = nucleotides. Dashed arrows indicate minor RNA polymerase III transcripts transcribed through the initial terminator sequences. (C) Restriction map of the hybrid plasmid. The black area is the 5S rRNA gene, the hatched area is plasmid sequences and the open area is rDNA. Restriction sites are: Xh, XhoI; St, SstII; Pv, PvuII; RI, EcoRI; Xm, XmnI; and (Dr), a DraI site used and destroyed in constructing the hybrid. T₅ and T₆ indicate TTTT and TTTT sequences respectively which terminate RNA polymerase III transcription. The short arrow indicates the site of cleavage. a, b, c and e indicate interpretations of the transcripts displayed in the corresponding tracks in part (B).

Figure 4: Snapback Transcripts Generated by an Inverted Repeat Template. (A) Map of the template. Restriction sites are: Hd, HindIII; Dr, DraI; Xh, XhoI; St, SstII; Ps, PstI. The region between the HindIII and PstI sites is duplicated. Triangles indicate the transcription initiation sites and the small arrows, the RNA cleavage sites. Production of specific double-stranded transcripts from the template is diagrammed, the long arrows representing the primary transcript. (B) Transcripts produced from the template. Reactions were standard and contained 1.2 $\mu\text{g/ml}$ of HindIII-cleaved (track a) or PstI-cleaved inverted repeat template DNA. Transcripts were treated with RNase prior to electrophoresis (see materials and methods). nt = nucleotides.

To confirm that this same cleavage event generated the 120 nt transcript we constructed a DNA template containing an inverted repeat of the rDNA 5' region. The structure of the template is shown in figure 4A. Transcription of this template should give rise to transcripts which rapidly snap back to form extensive regions of double stranded RNA which include the RNA processing site. It was hoped that rapid formation of the double stranded RNA (dsRNA) would preserve some of the primary transcripts from processing. Specific transcription was assayed by digestion of the snapback transcripts with RNase A to produce double-stranded RNA 'transcripts' with lengths corresponding to the distance from their original 5' end to the PstI site (diagrammed in figure 4A). If transcription normally commenced 120 bp upstream from the processing site, then the double-stranded versions of the primary transcript should be about 310 nt in length (120 nt plus 190 nt). Figure 4B, track a shows that in fact a prominent 320 nt dsRNA transcript was produced in this system. A prominent 200 nt transcript was also produced suggesting that some transcripts

were not protected from processing. Cleavage of the inverted repeat template with PstI, which should prevent the formation of snap-back transcripts, abolished formation of both the 320 nt and 200 nt transcripts (track b).

The results shown in figures 3B and 4B together indicate that the 120 nt transcript is cleaved from the 5' end of a precursor rDNA transcript. The lack of any higher molecular weight dsRNA transcripts in the experiment shown in figure 5B further suggests that the 5' end of the 320 nt transcript (and presumably the 120 nt transcript) represents the initiation site of rDNA transcription in vitro and is not generated by an additional processing event. This conclusion is supported by the experiment shown in figure 3B. In this experiment, the 5' fragment of the RNA polymerase III read-through transcript shows evidence of only one processing event, having a size of 520 nt. If two processing cleavages had occurred 120 nt apart, the 5' fragment would have been 400 nt long. No such 400 nt fragment is detectable, either in the experiment shown in figure 3B (track c) or using a transcription extract which simultaneously produced the 120 nt transcript as a result of RNA polymerase I transcription (data not shown). (The 410 nt transcript observed in tracks a and b in figure 3B derives from the 440 nt transcript not the 520 nt transcript, since it is absent from track c). These data therefore suggest that the 5' ends of the 120 nt and 320 nt transcripts define the initiation site of RNA polymerase I transcription in vitro.

In Vitro Transcription and Processing Accurately Reproduce In Vivo Events

To confirm that the 320 nt and 120 nt transcripts had the same 5' ends, and to determine if rDNA transcripts generated in vivo also shared this 5' end, the 5' ends of the in vitro transcripts and of rDNA transcripts from total Neurospora RNA were mapped by S_1 nuclease digestion of KNA.DNA hybrids. The 120 nt and 320 nt in vitro transcripts were purified by gel electrophoresis and hybridized to a RsaI-DraI DNA fragment 5' end-labelled within the transcribed region at the RsaI site (probe A in figure 5D). To rule out that hybridization was occurring to rDNA transcripts endogenous to the extracts, RNA of the same sizes as the in vitro transcripts was isolated from transcription reactions lacking template DNA (after digestion with RNase in the case of the 320 nt transcript) and then hybridized to the DNA probe. The RNA.DNA hybrids were then digested with S_1 nuclease. Figure 5A shows that although hybridization to the double stranded 320 nt transcript was very inefficient, the 320 nt transcript (track a) and the 120 nt transcript (tracks c and c') protected identical sets of DNA fragments of around 97 nt from S_1 nuclease, confirming that the two transcripts originated from the same site. In neither case was any hybridization to endogenous rDNA transcripts detected

Figure 5: Mapping the 5' Ends of In Vivo and In Vitro rDNA Transcripts. DNA probes singly labelled at the 5' position were hybridized to the relevant RNA preparation, then the RNA.DNA hybrids were digested with S_1 nuclease and resolved by acrylamide gel electrophoresis. All hybridizations contained 50 μ g *E. coli* RNA except those containing total *N. crassa* RNA. (A) *RsaI* probe (probe A). Hybridizations contained: track a, 320 nt dsRNA transcripts from 2 standard reactions; track b, 320 nt long, RNase-resistant endogenous transcripts from 2 reactions; tracks c and c', 120 nt 'run-off' transcript from 0.04 reactions; track d, 120 nt endogenous extract RNA from 0.04 reactions; track e, 50 μ g total *N. crassa* RNA; or, track f, 50 μ g total *E. coli* RNA. Tracks a-c were exposed 10 fold greater than tracks c'-f. (B) *SstII* probe (probe B). Hybridizations contained: track a, 50 μ g *E. coli* RNA; track b, 50 μ g total *N. crassa* RNA; track c, 320 nt RNase-resistant endogenous extract RNA from 2 reactions; track d, 320 nt dsRNA transcripts from 2 reactions. (C) *HpaI* probe (probe C). Hybridizations contained: track a, 200 nt dsRNA transcript from 2 reactions; track b, 200 nt RNase-resistant endogenous RNA from 2 reactions; track c, 190 nt *PstI* template run-off transcript from 0.04 reactions; track d, 190 nt endogenous RNA from 0.04 reactions; track e and e',

20 μ g total N. crassa RNA; track f, 20 μ g total E. coli RNA. Track e' was exposed five fold greater than tracks a-f. (D) Diagram of probes and the regions of them protected by the transcripts (solid line, protected; broken line, not protected). Asterisks indicate sites of 32 P labelling, the triangle indicates the initiation site and the arrow indicates the RNA cleavage site. (E) Sequencing strategy for the initiation region.

(tracks b and d). Further, rDNA transcripts from total Neurospora RNA protected precisely the same set of fragments around 97 nt in length (track e) as did the in vitro transcripts, suggesting that the initiation site for RNA polymerase I transcription is the same in vivo as in vitro.

To confirm that the 320 nt dsRNA transcript was generated by transcription through the processing site and to determine whether transcription proceeded through that site in vivo, the 320 nt transcript or total Neurospora RNA were hybridized to a SstII-DraI DNA fragment spanning the processing site and 5' end-labelled downstream at the SstII site (probe B in figure 5D). Figure 5B shows that the in vitro transcript (track b) and the Neurospora RNA (track d) protected an identical 165 nt DNA fragment from S_1 nuclease digestion. Since this represents the continuous distance from the SstII site to the transcription initiation site, spanning the processing site, these data confirm that RNA polymerase I transcription proceeds through the processing site both in vivo and in vitro. The localization of the initiation site 165 bp upstream of the SstII site also provides an explanation for a 165 nt transcript produced from SstII-cleaved templates (figure 3A, track b). This transcript must be the initial unprocessed run-off transcript of the SstII template, providing direct evidence for transcription through the processing site. Presumably termination of the transcript at the SstII site interferes with the normally very rapid and efficient cleavage reaction 40-45 nt upstream.

To determine if the 190 nt ssRNA transcript generated by run off at the PstI site (figure 1, track d), and the 200 nt dsRNA transcript generated from the inverted repeat template (figure 4B, track a) had the same cleavage-generated 5' end, and to determine if the cleavage event observed in vitro also occurred in vivo, the 190 nt or 200 nt in vitro transcripts, or total Neurospora RNA, were hybridized to a DNA fragment 5' end-labelled at a HpaI site 130 bp downstream from the processing site (probe C in figure 5D). Figure 5C (tracks a and c) shows that the two in vitro transcripts have identical 5' ends. Since both should also terminate at the same PstI site, the difference in mobility of the two transcripts must result from secondary structure differences or from differences in the processing of their 3'

Figure 6: Nucleotide Sequence of the RNA Polymerase I Promoter Region. Nucleotide +1 is the upstream-most of the two nucleotides at which transcription is presumed to commence. The other is +2. Restriction sites underlined were used to determine regions required for transcription in figure 8. Homologous sequences found in *N. crassa* α 5S rRNA genes (18) are indicated below the main sequence. The arrows indicate a region of dyad symmetry.

ends. Track e shows that rDNA transcripts exist in vivo with exactly the same 5' ends as the processed in vitro transcripts, indicating that the transcripts are processed at the same site in vivo as in vitro. Moreover, a longer exposure of track e (track e') reveals a longer 255 nt DNA fragment protected by unprocessed in vivo transcripts (indicated by the spot). The great disparity in abundance between the fragments protected by the processed and unprocessed transcripts (approximately 100-fold) indicates that processing at this site is highly efficient in vivo as well as in vitro.

Nucleotide Sequence of the RNA polymerase I Promoter Region

Figure 5D shows the sites of transcription initiation and RNA processing on a restriction enzyme map of the rDNA promoter region. In order to examine the DNA sequences involved in RNA polymerase I initiation, the nucleotide sequence of a 500 bp region surrounding the initiation site was determined according to the strategy shown in figure 5E. The sequence is shown in figure 6. In order to precisely locate the transcription initiation site within the nucleotide sequence, the 120 nt in vitro transcript was hybridized to a DNA fragment 5' end labelled at the KsaI site (probe A in figure 5D), digested with S_1 nuclease, and the protected DNA fragment co-electrophoresed with a chemical degradation sequencing ladder of the original KsaI fragment. The results, shown in figure 7, place the initiation site near the AccI site at two adjacent A residues designated +1 and +2 in figure 6.

Figure 7: Precise Localization of the Transcription Initiation Site. Probe A (figure 5D) was end-labelled at the *Rsa*I site, hybridized to purified 120 nt *in vitro* transcript from 0.04 standard transcription reactions and the KNA.DNA hybrids digested with S_1 nuclease. The protected DNA fragment was then electrophoresed alongside a chemical degradation (23) sequencing ladder of the original probe fragment. G, A, T and C indicate the products of the G-specific, A+G-specific, T+C-specific, and C-specific sequencing reactions respectively. Allowing for the different positions of cleavage by S_1 nuclease and the chemical sequencing reagents (37), the 5' ends of the transcript were localised to two adjacent A residues designated +1 and +2 in figure 6.

Sequences Required for RNA Polymerase I Transcription

As a preliminary step towards localizing the sequences required for RNA polymerase I transcription, restriction sites surrounding the initiation site were used to construct a series of plasmids containing 5' and 3' deletions of the promoter region. Figure 8A shows that a 5' deletion terminating at -96 at a DraI site does not affect transcription whereas 5' deletions terminating at -37 at a HaeIII site (track c) and at -17 at a XhoI site completely abolish transcription. Presumably a sequence between -96 and -37 is required for transcription. Similarly, figure 8B shows that 3' deletions to +165 (SstII site) and +95 (KsaI site) have no effect whereas a 3' deletion to -2 (AccI site) completely prevents transcription. Presumably a sequence between -2 and +95, perhaps the initiation site itself, is required for transcription.

DISCUSSION

We have previously described crude soluble extracts of Neurospora which can accurately initiate transcription of RNA polymerase III-dependant genes (20). We have shown here that under appropriate conditions, these extracts will also specifically initiate transcription of a cloned Neurospora rRNA gene. Transcription is insensitive to 1 mg/ml α -amanitin and to the presence of RNA polymerase III transcription-termination signals in the template DNA, and is therefore presumably carried out by RNA polymerase I. Under optimal conditions (very low DNA to extract ratios) the efficiency of transcription is 20-30 transcripts/gene/20 min. Higher DNA to extract ratios yield greater specific transcription but the relative efficiency is lower due to non-specific inhibition of the reaction by DNA.

Rapid processing accompanies transcription initiation

The 5' ends of the rDNA runoff transcripts lie 700 bp upstream of a XbaI site located near the 5' end of mature 17S rRNA. However, by constructing inverted repeat templates which generate snapback RNA transcripts resistant to processing, we have shown that the true initiation site of transcription lies at least a further 125 bp upstream, and that the 5' ends of the run-off transcripts are generated by a specific and very rapid cleavage reaction. The cleavage reaction was demonstrated directly by its effect on RNA polymerase III read-through transcripts initiated from a truncated 5S rRNA gene inserted upstream of the cleavage site. The cleavage reaction was so rapid that ordinarily no uncleaved transcripts could be detected in the run-off assays. The read-through transcripts were used to show further that the 5' end of the putative primary rDNA transcript is not generated by a second cleavage

Figure 8: Promoter Region Sequences Required for Transcription. (A) Transcription of 5' deletion templates. Subcloned templates were truncated at the *Dra*I site (-96), *Hae*III site (-37) or *Xho*I site (-17). WT indicates transcription of an undeleted (-1030) wild-type template. All template DNAs were cleaved at the *Eco*RI site and transcribed in standard reactions. (B) Transcription of 3' deletion templates. Subcloned templates (using pUC13 as vector) were truncated at the *Rsa*I site (+95) or the *Acc*I site (-2). Template DNAs were cleaved at the *Pst*I (WT, = +325) or *Sst*II sites (+165) or in flanking plasmid sequences at a *Pvu*II (+95) or *Nde*I site (-2). Transcription reactions were standard. The transcripts are all of the expected sizes. The doublet in the +95 track may result from 3' end processing.

reaction. Because the presence of phosphatases in the Neurospora transcription extracts precludes the identification of polyphosphate groups at the 5' termini of in vitro transcripts (20), we cannot rule out some exonucleolytic processing of the 5' end of the primary transcript. However, any such processing would have to be very rapid and specific since the double-stranded transcripts produced from the inverted repeat template had 5' termini identical to those of the run-off transcripts. Additional evidence that we have identified the true rDNA transcription initiation site is provided by the homology between the sequences surrounding this site and the S. carlsbergensis initiation site (27) (14 out of 20 bp) (see below).

In vivo transcripts from total Neurospora RNA have 5' ends corresponding precisely to those of both the primary and processed in vitro transcripts. No other in vivo rDNA transcripts were detected, even after long exposure of the autoradiographs. Hence the RNA polymerase I initiation site identified in vitro appears to be the site utilized in vivo. Similarly, the cleavage event occurring 120-125 nt from the initiation site in vitro appears to also occur in vivo with a similar efficiency and identical specificity.

Conserved elements in the Neurospora RNA polymerase I promoter region

We have determined the nucleotide sequence of the 400 bp region surrounding the RNA polymerase I initiation site and have shown that sequences between -96 and +97 are sufficient for full transcription in an in vitro run-off assay. The nucleotide sequence shows little homology to that surrounding the rDNA initiation site of yeast, a related ascomycete, with the exception of the region from -16 to +3 which displays 14 matches with the S. cerevisiae and S. carlsbergensis sequences (27) and 12 matches with the S. rosei sequence (27). Indeed, as shown in figure 9, the sequence surrounding the Neurospora initiation site (-16 to +9) displays a degree of homology with the initiation regions of the nine organisms for which sequences have been reported, including yeast, slime moulds, Drosophila and humans. The most convincing homologies are with Dictyostelium discoideum (13/14 matches between -5 and +9), Tetrahymena pyriformis (11/14 between -5 and +9), rat (11/15 between -14 and -2), and human (11/15 between -15 and -2). The most marginal similarities are with Xenopus spp. and Physarum polycephalum. In particular, the sequence TATA, or similar, occurs between -5 and +1 in 9 of the 10 organisms, the exception being Xenopus spp. The partial conservation of the initiation region of these very disparate organisms is matched by a strong conservation of the region between more closely related organisms (33). The possible importance of this region for transcription initiation is also suggested in

Figure 9: Homology of the *Neurospora* Initiation Region to the Initiation Regions of Other Organisms. Dashes indicate identical nucleotides. Parentheses and slashes indicate arbitrary deletions and insertions, respectively, introduced to maximize homology to the *Neurospora* sequence. Dots indicate the start points of transcription. Sequences beyond those indicated show no homology at all to the corresponding *Neurospora* sequences. The boxes indicate a TATA-like element partially conserved among 9 of the 10 organisms. Sequences are from ref. 27 (*S. carlsbergensis*), ref. 28 (human, rat and mouse), ref. 29 (*X. borealis*), ref. 10 (*D. melanogaster*), ref. 30 (*T. pyriformis*), ref. 31 (*P. polycephalum*) and ref. 32 (*D. discoideum*).

Neurospora by the fact that 3' deletions to -2 abolish *in vitro* transcription (figure 8B). Similarly, in mouse, 3' deletions to -2 reduce transcription 50-fold (11).

The homology of the *Neurospora* initiation region to those of the other organisms, including yeast, is strictly confined between -16 and +9, with one exception. In the case of *Dictyostelium discoideum*, the region of homology extends downstream to +47 and there is an additional 17 bp of homology upstream of -5 (figure 9). The strong conservation of this long region (74% homology over 50 bp from -3 to +47) is particularly remarkable considering the divergence of the two species and the disparity of the yeast and *Neurospora* sequences. For example, whereas the yeast and *Neurospora* 5.8S rRNAs are 92% homologous, those of *Neurospora* and *Dictyostelium* are only 51% homologous

(35). The similarity of this extended region in Neurospora and Dictyostelium may be a chance evolutionary relic or may indicate that a particular function has been retained for the sequence by these two species. This hypothetical function could relate to transcription initiation but alternatively might involve the 120 nt RNA cleaved from the 5' end of the Neurospora primary transcript.

Sequences Required for Transcription Initiation

We have shown that sequences required for RNA polymerase I transcription in Neurospora lies between -96 and -36 (figure 8A). One striking feature of this region is a long (23 bp) region of dyad symmetry between -95 and -73, indicated by the arrows in figure 6. A similar region (-99 to -79) is strongly conserved between S. rosei and S. carlsbergensis, but does not display dyad symmetry in these organisms (27). In comparison, in Drosophila (10), Xenopus (36) and mouse (9,11), sequences upstream of -43 are not directly required for full transcription in vitro. However, competition (9) and in vivo assays (36) suggest that upstream sequences can play an auxiliary role.

Common sequences in the Neurospora rRNA and 5S rRNA promoter regions

Since we were interested in the relationship between transcription of the large rRNA and 5S rRNA genes, we compared the sequence of the 5S rRNA gene and surrounding region (18) with that of the large rRNA promoter region. Two significant homologies were observed. The sequences, -62-GAAAAGTCG and -25-CCAGACTCG in the large rRNA promoter closely resemble the 5S rDNA sequences, +21-GAAAAGTCG and +68-CCAGACTAG respectively, displaying 8/9 matches in each case (figure 6). The precise biological significance of these homologies is uncertain. In the large rRNA promoter, -62-GAAAAGTCG occurs with the region required for transcription, and a similar region (-61 to -56) is conserved between yeast species (27). Moreover, the 5S rDNA counterpart is required for 5S rRNA transcription (B. Tyler, unpublished). Similarly, +68-CCAGACTAG constitutes part of another element of the Neurospora 5S rDNA promoter (B. Tyler, unpublished). Its large rDNA counterpart, between -25 and -17 has not yet been implicated in transcription in Neurospora but a similar region in mouse (-26 to -17) is required for minimal levels of transcription in vitro (11) and contains a sequence conserved in humans (28). The occurrence of these two sequences in both the large rDNA and the 5S rDNA promoter regions raises the possibility that in Neurospora there may be direct co-ordination of the transcription of these two gene families.

ACKNOWLEDGEMENTS

We thank R. Metzenberg and S. Free for a gift of plasmid pKD2, and C. Todd and K. Gardner for excellent technical assistance. B.M.T. was the recipient of a Damon Runyon-Walter Winchell Cancer Fund Fellowship, DRG-624. This work was supported by N.I.H. grant GM28777 (to N.H.G.).

*To whom reprint requests should be sent. Present address: Department of Genetics, Research School of Biological Sciences, Australian National University, P.O. Box 475, Canberra City, ACT 2601, Australia

REFERENCES

1. Swanson ME and Holland MJ (1983) *J Biol Chem* **258**, 3242-3250
2. Kohorn BD and Rae PMM (1982) *Proc Nat Acad Sci USA* **79**, 1501-1505
3. Wilkinson JK and Sollner-Webb B (1982) *J Biol Chem* **257**, 14375-14383
4. Grummt I (1981) *Proc Nat Acad Sci USA* **78**, 727-731
5. Mishima Y, Yamamoto O, Kominami R and Muramatsu M (1981) *Nucl Acids Res* **9**, 6773-6785
6. Miller KG and Sollner-Webb B (1981) *Cell* **27**, 165-174
7. Mishima Y, Financsek L, Kominami R and Muramatsu M (1982) *Nucl Acids Res* **10**, 6659-6670
8. Miesfeld K and Arnheim N (1984) *Mol Cell Biol* **4**, 221-227
9. Grummt I (1982) *Proc Nat Acad Sci USA* **79**, 6908-6911
10. Kohorn BD and Rae PMM (1983) *Proc Nat Acad Sci USA* **80**, 3265-3268
11. Yamamoto O, Takakusa N, Mishima Y, Kominami R and Muramatsu M (1984) *Proc Nat Acad Sci USA* **81**, 299-303
12. Segall J, Matsui T and Roeder KG (1980) *J Biol Chem* **255**, 11986-11991
13. Matsui T, Segall, J, Weil PA and Roeder KG (1980) *J Biol Chem* **255**, 11992-11996
14. Sakonju S, Bogenhagen DF and Brown DD (1980) *Cell* **19**, 13-25
15. Bogenhagen DF, Sakonju S and Brown DD (1980) *Cell* **19**, 27-35
16. Corden J, Wasylyk B, Buchwalder A, Sassone-Corsi P, Kedinger C and Chambon P (1980) *Science* **209**, 1406-1414
17. Free SJ, Rice FW and Metzenberg RL (1979) *J Bact* **137**, 1219-1226
18. Selker EU, Yanofsky C, Driftmier K, Metzenberg RL, Alzner-DeWeerd B and Raj Bhandary UL (1981) *Cell* **24**, 819-828
19. Tyler BM and Giles NH (1985) *Proc Nat Acad Sci USA* **82**, in press
20. Tyler BM and Giles NH (1984) *Nucl Acids Res* **12**, 5737-5755
21. Messing J (1983) *Methods in Enzymology* **101**, 20-78
22. Tyler BM, Geever RF, Case ME and Giles NH (1984) *Cell* **36**, 493-502
23. Maxam AM and Gilbert W (1980) *Methods in Enzymology* **65**, 499-560
24. Horowitz B, Goldfinger BA and Marmur J (1976) *Arch Biochem Biophys* **172**, 143-148
25. Bogenhagen DF and Brown DD (1981) *Cell* **24**, 261-270
26. Dignam JD, Lebovitz RM and Roeder RG (1983) *Nucl Acids Res* **11**, 1475-1489
27. Verbeet MPH, Klootwijk J, Heerikhuizen Hv, Fontijn R, Vreugdenhil E and Planta RJ (1983) *Gene* **23**, 53-63
28. Financsek I, Mizumoto K, Mishima Y and Muramatsu, M (1982) *Proc Nat Acad Sci USA* **79**, 3092-3096
29. Furlong JC, Forbes J, Robertson M and Maden BEH (1983) *Nucl Acids Res* **11**, 8183-8196
30. Saiga H, Mizumoto K, Matsui T and Higashinakagawa T (1982) *Nucl Acids Res* **10**, 4223-4236

31. Blum B, Seebeck T, Braun R, Ferris P and Vogt V (1983) Nucl Acids Res 11, 8519-8533
32. Hoshikawa Y, Iida Y and Iwabuchi M (1983) Nucl Acids Res 11, 1725-1734
33. Sommerville J (1984) Nature 310, 189-190
34. Grummt I, Koth E and Paule MK (1982) Nature 296, 173-174
35. Erdmann VA, Wolters J, Huysmans E, Van den Berghe A and Wachter RDe (1984) Nucl Acids Res 12, r133-r166
36. Keeder KH, Roan JG and Dunaway M (1983) Cell 35, 449-456
37. Sollner-Webb B and Keeder KH (1979) Cell 18, 485-499